ARABESQUE DATABASE CONTENTS

Introduction

We first of all need to clarify what is meant by the 'Arabesque Database' in the context of this document. It is perfectly valid to describe the whole of the CD-ROM as the 'database', including the manual which includes material such as images that are data in their own right.

However, a significant subset of the ARABESQUE data has been organised into an integrated structure within a relational database. It is to these data that this document refers.

The database contains a vast amount of data. In terms of sheer volume it occupies some 10 Mbytes in a relatively compact format. This doubles when the overheads associated with storage in a relational database management system are added. However, it is not the volume of the database that is its most impressive statistic, but the diversity of the oceanographic data it contains.

The data model underlying the database is based on the concept of events. An event is defined as an action that results in the generation of oceanographic data. Events can be many things ranging from the spectacular such as a mooring deployment to the trivial such as turning on a tap to collect a sample from the non-toxic supply.

Broadly speaking, events are of two types. Point events are events that may be considered to relate to a single position, like CTD profiles, vertical net hauls or corer deployments. Traverse events are events that occur along a significant distance and generally pertain to oceanographic hardware that is towed by a ship.

However, it should be noted that the division between point and traverse events is not clear cut depending upon the scale at which the events are viewed. During a CTD cast, the ship is never completely still and at high resolution the CTD event could be considered as a traverse event. Likewise, an instrument tow on a tight grid within a small box could be considered a point event.

Within the database, event types are assigned on the basis of common sense with a view to making the interpretation of the data as easy as possible. This inevitably results in events being classified as point events wherever possible.

For the purposes of this document, the data within the database are grouped into broad categories designed with the way in which data are viewed by the oceanographic community very much in mind.

The following groupings are used:

Event Inventory
Data Dictionaries
Water Column Profile Data
Rate Measurements

As with any classification, there are grey areas where groups overlap or where the groupings based on oceanographic criteria do not map cleanly with optimum data storage structures. Care has been taken to identify these problem areas in the text. The objective of this document is to provide an overview of the different types of data held in the database and to provide guidance on where to look for them.

Event Inventory

The Event Inventory is probably the most important component of the database. In addition to its inventory function, it stores data on event attributes, such as space and time co-ordinates. Obviously, without this information the data in the database would be useless.

The bulk of the inventory information is held in the EVENT table which contains times, positions and other ancillary information. This is supported by table EVENT_COMM which provides storage for plain language comments and G_CODE which defines mnemonics used in table EVENT. The table CRSINDX provides additional information on the cruises associated with the events.

In addition to this primary index of events, there is also a series of secondary indices that contain additional information that is specific to a particular type of oceanographic data. This information could have been stored in EVENT, but it would have resulted in an unmanageable number of columns. However, these data are often of equal importance to the data stored in EVENT. For example sample depths are stored in table BOTTLE and without these, sample data would be useless.

The following secondary indices are present in the database.

Water and air samples BOTTLE
CTD data CTDINDX, CTDCAL
Radiometer profiles PRINDX
Instrument type codes used in table CTDINDX are defined in table CTDTYP.

Data Dictionaries

The major data tables in the database use coded fields to store information. The most important of these fields are those identifying the parameter measured and the data originator.

The coding convention used for parameters is defined by a group of tables known as the Parameter Dictionary, namely ZUSG, ZUPM and ZUNT. The bulk of the code definition is stored in ZUSG, including the parameter name and the protocol used to measure it. The units in which the parameter is stored may be found by obtaining a coded field from table ZUPM that is defined in table ZUNT.

There is another table in the Parameter Dictionary called ZUCT but this is only of use to those wishing to assign parameter codes to data (i.e. BODC personnel).

Data originators are identified by simple numeric codes that are defined in table ORGCODE.

Water Column Profile Data

There are several different types of water column profile data held in the database. Broadly, these fall into two different types namely sample profiles and instrumental profiles.

Sample profiles, often known as 'bottle data', may be found in table BOTDATA. Many different parameters are stored here and as a general rule if any type of measurement made on a discrete water sample is required then this is the place to look.

BOTDATA has a highly normalised structure with a record for each parameter measured on each sample by a given originator. The possibility (and actuality) that more than one originator might measure the same parameter on the same sample gives rise to a nasty complication (a straightforward cross-tabulated query fails) when retrieving the data. Because of this, it is recommended that retrieval of the data from BOTDATA should only be attempted using the Database Explorer software.

The database contains several types of instrumental water column profile. Profiling radiometer data may be found in table PRPROF. However, there are many CTD casts in the database which have downwelling and upwelling irradiance channels included. These are not included in PRPROF and must be obtained from table BINCTD.

Table BINCTD is predominantly concerned with the storage of CTD profiles. The data set includes all the parameters measured by the CTD package.

Rate Measurements

The rate measurements in the database are measurements of uptake kinetics by the particulate phase from the dissolved phase. These consist of what are broadly termed production experiments.

Wherever possible, rate measurements have been treated as water sample parameters and are stored in table BOTDATA. Thymidine and leucine incorporation data will be found here, for example.

BOTDATA has a normalised structure which is complicated by the fact that different originators have measured the same parameter on the same sample. It is therefore recommended that it be interrogated using the Database Explorer software.

In a number of cases, such as 24 hour incubation it has proved impossible to store the data in BOTDATA with sufficient information for them to be of use. In these cases, bespoke table structures have been used.

The following tables are involved:

Oxygen/ TCO_2 production experiments OXYHDR and OXYDAT.